DOCUMENT RESUME

ED 198 485 CS 005 874

AUTHOR Raygor, Alton L.; Flippo, Rona F.
TITLE Varieties of Comprehension Measures: A Comparison of Intercorrelations among Several Reading Tests.
PUB DATE Dec 80
NOTE 17p.; Paper presented at the Annual Meeting of the American Reading Conference (1st, Sarasota, FL, December 4-6, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Higher Education; Postsecondary Education; Predictor Variables; *Reading Achievement; *Reading Comprehension; Reading Research; *Reading Tests; Secondary Education; *Test Validity
IDENTIFIERS *Minnesota Reading Assessment

ABSTRACT

The procedures used to establish the validity of the Minnesota Reading Assessment (MRA) are described in this paper. The MRA is described as a measure designed for use with students in community colleges, business schools, technical schools, and other secondary and postsecondary training institutions in order to measure student competence in specific reading skills: reading rate and retention, vocabulary, and comprehension. The paper presents data drawn from the validity studies, which were completed in connection with the standardization of the test and which involved students in vocationally oriented programs, in community colleges, in four-year colleges, and in programs for the hearing impaired. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

VARIETIES OF COMPREHENSION MEASURES: A COMPARISON OF INTERCORRELATIONS AMONG SEVERAL READING TESTS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Alton L. Raygor - University of Minnesota
Rona F. Flippo - Georgia State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alton L. Raygor
Rona F. Flippo
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the First Annual Meeting of the American Reading Conference, Sarasota, Florida, December, 1980

ED198485

The purpose of this paper is twofold: First, to discuss some measurement problems in reading comprehension, and second, to report some research results on a new reading test.

Of all the terms used, and variables presumably measured in the testing of reading performance, none is more difficult to define or specify than the construct, "reading comprehension". A quick consideration of the other two most frequently measured variables, rate and vocabulary, shows real differences in the clarity of what we are measuring.

In the case of reading "rate", we have a nice, neat, numerical concept to lean on. It is the number of words read per minute. That sounds very simple and straightforward, and if we ignore the problems of reliability and validity, it is. At least we know what is being measured, even if some of us are not convinced that it is the most important thing to measure.

When we talk about measuring "vocabulary", we can say some very general, widely accepted things about it. Vocabulary tests are usually thought of as reasonably reliable, and probably valid if they have face validity or content validity for the particular student and curriculum. There are problems when we try to specify exactly what we mean by "knowing a word", but most of us do not reject the test because of these difficulties.

CS205874

Comprehension, on the other hand, is a much fuzzier notion, and efforts to define it or explain it are less likely to be satisfactory. Efforts to demonstrate types of comprehension, for example, have produced sub-measures that typically do not hold up well in empirical studies. The scores on comprehension sub-scales typically intercorrelate at about the level of their reliability, leading us to question whether they are all really measuring the same thing.

Recent efforts to understand comprehension by means of an analysis of the syntax show promise in understanding the nature of language, and perhaps will lead to better understanding of the comprehension process, but do not solve our measurement problems yet.

In a sense, we are measuring comprehension in all parts of all reading tests, if we consider the underlying process and use the word comprehension in its general meaning to understand. Presumably the rate measure gives us some notion of the speed with which the reader understands the material. Retention measures also presume that the reader has understood the material and has retained some facts. Few will disagree that vocabulary knowledge is necessary for the comprehension of reading material.

What, then, are we measuring when we present the student with the usual comprehension test consisting of a set of short passages followed by questions? And how does the result of such a performance relate to the other aspects of the reading process? We are not prepared to give a definitive answer to the first question concerning what we are really measuring. However, it seems clear that in some sense reading comprehension occurs at several levels, and we believe that the typical reading comprehension test gets at a higher-order set of behaviors than simple literal understanding and retention of detail.

The student is asked questions concerning main ideas, relationships among ideas, conclusions that can be drawn from the passages, and in some cases is asked to deal with the author's purpose and general approach.

As the questions require more and more manipulation and interpretation of the information given, they seem to become more involved with a reasoning process in addition to the passive reception of the author's message. At some point it is difficult to say whether we are measuring reading ability or general verbal reasoning ability.

As we turn to the other question concerning relationships among various reading test measures, we are fortunately able to examine some empirical data. The data to be presented are the result of several concurrent validity studies done in connection with the standardization of a new reading test, the Minnesota Reading Assessment (MRA).

The MRA is designed for administration to students in community colleges, business schools, technical schools, and other secondary and post-secondary training institutions. It is intended to measure the students' competence in those reading skills which are most relevant to school success. The test, available in two forms, A and B, measures Reading Rate and Retention, Vocabulary, and Paragraph Comprehension. Working time is forty minutes.

The MRA was developed to provide information that would identify students who are in need of remedial assistance. The stimulus materials in the test are especially selected so as to be real-world oriented, vocationally relevant content. The difficulty level is such that the test will discriminate best on the students in the lower portions of the distribution, where judgments about remediation are usually made.

Normative information is provided on (1) Student in Vocationally Oriented Programs, (2) Students in Community Colleges, (3) Students in Four-Year Colleges, and (4) Hearing Impaired Students.

Reliability Studies

In a study of the internal characteristics of the instrument at the University of South Carolina's College of General Studies, Columbia, ninety-five students were studied. Fifty-one took Form A and forty-four took Form B. Reliability estimates resulting from that study are shown below in Table 1.

Table 1 - Reliability

	Form A	Form B
Cronbach's Alpha	.917	.934
Horst Modified	.924	.941
Kuder-Richardson 20	.917	.934
Kuder-Richardson 21	.900	.920

In another study, interform reliability was studied using data from a sample of 24 students at the University of Wisconsin-Stout, in Menomenie, Wisconsin. The resulting correlations are shown in Table 2.

Table 2 - Interform Reliability

	Rate	Retention	Vocabulary	Paragraph Comprehension	Total Score
Correlations between Form A and Form B N=24	.76	.53	.83	.57	.82

Intercorrelation of Part Scores

In a study of one hundred one freshmen students attending the University of South Carolina's College of General Studies in 1980, fifty-three students took Form A and forty-eight students took Form B of the MRA. The part-score intercorrelations are shown in Table 3.

Table 3 - MRA Scale Intercorrelations

Sample: 101 Students
University of South Carolina, College of General Studies
Developmental Center

	Form A N=53					Form B N=48				
	Rate	Retention	Vocab.	Para. Comp.	Total	Rate	Retention	Vocab.	Para. Comp.	Total
Rate	1.00	.09	.37	.20	.34	1.00	.32	.41	.35	.46
Ret.		1.00	.20	.40	.54		1.00	.41	.52	.67
Voc.			1.00	.52	.91			1.00	.41	.92
P.C.				1.00	.75				1.00	.67
Total					1.00					1.00

During the process of standardizing the test, it was possible to assemble data regarding performance on other reading tests on the part of the students in the standardization sample. Such concurrent validity data form the substance of this report.

Studies were done at the University of South Carolina's College of General Studies' Developmental Center in Spring 1980 to determine the relationships between scores on the Minnesota Reading Assessment and two other standardized tests, the Nelson-Denny Reading Test, Form C and Form D, and the Stanford Diagnostic Reading Test, Blue Level. We were interested in seeing the correlations between these instruments and the MRA, and were also interested in the relative difficulty of the tests.

The population for these studies consisted of freshmen students mandatorily enrolled in effective reading classes due to SAT verbal scores below 390 and freshmen students referred to the Center for reading or study skills assistance because they were observed by faculty to be having academic problems.

The Nelson-Denny was selected as one of the comparison tests because surveys in community college reading and reviews of the literature have indicated that the Nelson-Denny is the most widely used reading test for that population (Goodwin, 1971; Sweiger, 1972; Landsman and Cranney, 1978; Flippo, 1980a).

The Stanford Diagnostic, Blue Level, was selected because it was developed to meet the reading assessment needs of grade 9 through junior/community college students, particularly of low-achieving students, and has norms for the junior/community college group. Van Roekel (1978) in The Eighth Mental Measurements Yearbook stated "The SDRT has few peers among group diagnostic reading tests" (p. 1299).

Correlations With Other Reading Tests

Criterion-Related Validity

In the University of South Carolina study, comparing the MRA with the Nelson-Denny Reading Test to determine the relationships between scores, fifty-three students took Form A of the MRA, and forty-eight students took Form B. All students had previously taken the Nelson-Denny Reading Test. The following table shows the Pearson Product-Moment Correlations between the two tests:

Table 4

Correlations Between the Minnesota Reading Assessment and the Nelson-Denny Reading Test

Nelson-Denny Reading Test	Minnesota Reading Assessment									
	Form A N=53					Form B N=48				
	Rate	Retention	Vocab.	Para. Comp.	Total	Rate	Retention	Vocab.	Para. Comp.	Total
Rate	.42	-.11	-.07	.01	-.08	.39	-.17	.10	-.01	.03
Vocabulary	.34	.11	.39	.37	.41	.47	.18	.54	.15	.48
Comprehension	.21	.24	.21	.43	.33	.42	.45	.51	.50	,61
Total Score	.31	.21	.34	.47	.42	.52	.36	.61	.37	.62

In the related study at the University of South Carolina eighty-nine students took both the MRA and the Stanford Diagnostic Reading Test, Blue Level. Forty-four students took Form A and forty-five students took Form B of the MRA. Table 5 shows the Pearson Product-Moment Correlations between the two tests.

Table 5 inserted here.

In a different study at the University of Minnesota, one hundred twenty-two students took both Form A of the McGraw-Hill Basic Skills System Reading Test and the MRA. The Pearson Product-Moment Correlations between the total scores on the two were as follows:

	N	r
Form A	59	.72
Form B	63	.82

In another study at Lakewood Community College, seventeen students took

Table 5

Correlations Between the Minnesota Reading Assessment and the Stanford Diagnostic Reading Test, Blue Level

Stanford Diagnostic Reading Test	Minnesota Reading Assessment									
	Form A N=44					Form B N=43				
	Rate	Retention	Vocab.	Para. Comp.	Total	Rate	Retention	Vocab.	Para. Comp.	Total
Literal Comp. LC	.32	.24	.29	.37	.40	.56	.60	.60	.48	.63
Inferential Comp. IC	.44	.08	.27	.54	.35	.48	.38	.51	.43	.57
Word Meaning WM	.18	.24	.51	.53	.57	.49	.49	.74	.54	.80
Word Parts WP	.08	.01	.29	.20	.26	.20	.38	.42	.40	.50
Phonetic Analysis PA	-.09	.10	.20	.33	.26	.25	.25	.39	.39	.44
Structural Analysis SA	-.01	.05	-.02	.09	.03	.04	.38	.31	.48	.44
Scanning & Skimming SS	.03	.19	.08	.24	.06	.43	.39	.41	.38	.49
Fast Reading FR	.07	.25	-.04	.08	.08	.19	.28	.21	.11	.23

Form A of the MRA and Form C of the Nelson-Denny Reading Test. The correlation between total scores was .79.

Relative Item Success Study

One way of looking at the appropriateness of a reading test for a given population can be students' ability to finish at least half the test correctly. As part of the University of South Carolina's study, we looked at the number of correct on each subtest of the Stanford Diagnostic, Blue Level, the Minnesota and the Nelson-Denny. Students were grouped as "correctly answering less than half of the test items" (49% correct or less) or "correctly answering half or more than half of the test items" (50% correct or more). Table 6 shows the data gathered from this study of relative item success on the three tests.

Table 6 inserted here.

Results

On the subtests of the Stanford Diagnostic, more students were successful at completing half or more than half of the test items on each subtest. However, there were enough that could not answer half of the items correctly, indicating that this test does discriminate well for lower-achieving freshmen students and for those in somewhat higher ability groups in college reading improvement situations.

On the subtests of the Minnesota, most of the student could successfully complete half of the items. (This could lead to the implication that the Minnesota might be considered "easy" for these students.)

On the subtests of the Nelson-Denny, most of the students could not be successful on half of the items, especially on the vocabulary items.

Table 6

Relative Item Success on the Stanford Diagnostic Reading Test - Blue Level, Minnesota Reading Assessment, and the Nelson-Denny Reading Test - Forms C and D

Population: Freshmen Students Pretested at University of South Carolina's College of General Studies Developmental Center, Spring Semester 1980

	Literal Comprehension (SDRT)	Inferential Comprehension (SDRT)	Word Meaning (SDRT)	Word Parts (SDRT)	Phonetic Analysis (SDRT)
Total Number of Students	187	187	187	187	187
Correctly Answered less than half of the test items	47	86	81	80	98
Correctly Answered half of the test items or more	80	101	106	107	89

SDRT = Stanford Diagnostic Reading Test, Blue Level
MRA = Minnesota Reading Assessment
NDRT = Nelson-Denny Reading Test, Forms C and D

Table 6 (continued)

	Structural Analysis (SDRT)	Scanning & Skimming (SDRT)	Fast Reading (SDRT)	Retention (MRA)	Vocabulary (MRA)
Total Number of Students	127	127	127	114	114
Correctly Answered less than half of the test items	21	40	64	16	16
Correctly Answered half of the test items or more	106	87	63	98	98

	Paragraph Comprehension (MRA)	Vocabulary (NDRT)	Comprehension (NDRT)
Total Number of Students	114	127	127
Correctly Answered less than half of the test items	9	114	84
Correctly Answered half of the test items or more	105	13	43

SDRT Stanford Diagnotic Reading Test, Blue Level
MRA - Minnesota Reading Assessment
NDRT - Nelson_Denny Reading Test, Forms C and D

Implications

The Stanford Diagnostic is a good group diagnostic test and does discriminate well for this population. We are not surprised by its item success distributions, since a good diagnostic test for this population should look like this. We suggest that the Stanford Diagnostic, Blue Level, is one test that could adequately be used to get a more diagnostic picture of each students' reading abilities and inabilities prior to prescription in college reading improvement programs for the less than adequately prepared students. (Also see findings in Flippo, 1980b for more information on SDRT for diagnosis.)

The Minnesota and Nelson-Denny are not as diagnostic as the Stanford. They are survey tests and are designed to give the screener a general estimate of where a student stands. Is s/he in need of help or not in need of help? This should be asked if the student is being tested to see if s/he requires reading skills assistance. Or is s/he an excellent reader or not very excellent reader? This should be asked if the student is being tested to see if s/he should be matriculated into a special course or program where reading excellence can screen one in or out.

From our observations, the MRA has low item difficulty and identifies students having real problems with weak skills. These are students in need of reading remediation in order to have a chance to succeed in a community college program. Tittle and Kay, in a paper presented to the American Educational Research Association in 1971, called for a test that could adequately identify the lower half of the achievement distribution. The MRA can do that.

From our observations the NDRT is much too difficult for this population and is not an appropriate test to identify students who need help with reading skills in a less than adequately prepared group. It undoubtedly causes frustration

and loss of esteem to students who are already frustrated by their academic abilities. The NDRT has a very high item difficulty level and is excellent for identifying highly skilled readers.

Discussion

Some interesting relationships appear. For one thing, in spite of extreme efforts to make Form A and Form B of the MRA as parallel and equivalent as possible, some real differences appeared in comparisons with other tests. Form B is consistently more highly correlated with the Stanford, the McGraw Hill, and the Nelson-Denny.

The vocabulary measure also are interesting. The various scores labeled "vocabulary" correlate with the MRA vocabulary score at moderate levels, ranging from .39 to .74. However, the vocabulary scores on other tests correlate with the MRA total score at consistently higher levels, ranging from .41 to .80.

Measures labeled "comprehension", in the other hand, correlate with MRA Paragraph Comprehension from .43 to .60 (X = .51). They correlate with MRA Total score from .33 to .61 (X = .48). This pattern, even though the differences are not great, shows a reversal of the pattern on Vocabulary scores.

We do not feel that there are any very surprising results in any of this, except possibly the high correlations between Stanford comprehension and MRA Rate. One purpose of the studies was to examine the concurrent validity of the MRA. We conclude that such validity has been demonstrated at about the level usually shown by studies of nationally standardized tests.

As a result of the relative item success study, we found that the MRA is quite easy for the freshmen population in a reading improvement program at a major state university. (It should be pointed out that the MRA was designed

for the technical and community college student population. However, we feel that "easy" is to be preferred to "difficult" when the screener is interested in testing to see who has real reading skills problem and just how weak these skills may be.

The relative item success study suggests a need for yet another new survey instrument that could adequately measure students in college reading improvement programs reading above the MRA level but below the Nelson-Denny (probably in the ninth to twelfth grade range). A test like that would show distributions similar to those shown by the SDRT with the same population.

Finally, before one decides between the NDRT or MRA or any other reading test, one must know why it is being given and to what population. One should also consult a reliable source, such as a current Buros to see what reviewers say about the difficulty of test items on all instruments being considered. With that information in mind, we think that a good selection of tests can be made. If you want to see who has weak skills, use something that discriminates among low ability students. If you want to see who your very good readers are, use something that discriminates among higher ability students.

We do not think our studies have shed much light on the nature of "comprehension", but we do feel that we have supported the need for appropriate tests to measure reading comprehension for different ability groups.

References

Brown, J.I., Nelson, M.J. and Denny, E.C., The Nelson-Denny Reading Test, Forms C and D. Boston: Houghton Mifflin Company, 1973.

Flippo, R.F. Diagnosis and prescription of college students in developmental reading programs - a review of the literature. Reading Improvement, 1980, 17, (4) (a)

Flippo, R.F. The need for comparison studies of college students' reading gains in developmental reading programs using general and specific levels of diagnosis. In M.I. Kamil and A.J. Moe (Eds.), Perspectives on reading research and instruction. Twenty-ninth Yearbook of the National Reading Conference. Washington, D.C.: National Reading Conference, 1980. (b)

Goodwin, D.D. Measurement and evaluation in junior college reading programs. Junior College Research Review, October 1971, 6 (2), 1-3. (ERIC Document Reproduction Service No. ED 053 714)

Karlsen, B., Madden, R. and Gardner, E.F. Stanford Diagnostic Reading Test, Blue Level, Form A and Form B. New York: Harcourt Brace Jovanovich, 1976.

Landsman, M.B. and Cranney, A.G. Training and activities of Florida community college reading teachers. Florida Reading Quarterly, 1978, 14, (2), 17-22.

Raygor, Alton L. Manual for the Minnesota Reading Assessment. Rehoboth, Massachusetts: Twin Oaks Publishing, Inc., 1980.

Spaulding, J.D. Designs and organizational structures of junior and community college reading programs across the country. In F.B. Greene (Ed.), College reading: Problems and programs of junior and senior colleges (Vol. 2). The 6th Yearbook of the National Reading Conference. Boone, N.C.: National Reading Conference, 1972.

Tittle, C. and Kay, P. Selecting tests for an open admissions population. Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971. (ERIC Document Reproduction Service No. 048 359)

Van Roekel, B.H. Stanford Diagnostic Reading Test. In O.K. Buros (Ed.) The eighth mental measurements yearbook (Vol. 2). Highland Park, New Jersey: The Gryphon Press, 1978, 1297-1300.